

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

На правах рукопису  
УДК 519.254

До захисту допущено  
В. о. завідувача кафедри ММСА

О.Л. Тимошук

« \_\_\_\_ » \_\_\_\_\_ 2020 р.

## Магістерська дисертація

на здобуття ступеня магістра за спеціальністю 124 Системний аналіз  
на тему: «Система максимізації вартості клієнта (МВК) із використанням  
нейронних мереж у сфері роздрібної торгівлі»

Виконав:

студент II курсу, групи КА-92 мп  
Свириденко Віталій Олександрович

\_\_\_\_\_

Керівник:

професор кафедри ММСА,  
д.ф.-м.н., проф. Макаренко О.С.

\_\_\_\_\_

Рецензент:

молодший науковий співробітник  
ННК "ІПСА" НАН України,  
к.т.н. Осауленко В.М.

\_\_\_\_\_

Засвідчую, що у цій магістерській дисертації  
немає запозичень з праць інших авторів  
без відповідних посилань  
Студент \_\_\_\_\_

Київ  
2020

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ СІКОРСЬКОГО»  
ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ  
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

Рівень вищої освіти — другий (магістерський)  
Спеціальність — 124 «Системний аналіз»

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри ММСА

О. Л. Тимощук

«\_\_\_» \_\_\_\_\_ 2020 р.

**ЗАВДАННЯ**

на магістерську дисертацію студенту Свириденко Віталію Олександровичу

**1. Тема дисертації:** «Система максимізації вартості клієнта (МВК) із використанням нейронних мереж у сфері роздрібної торгівлі», науковий керівник дисертації Макаренко Олександр Сергійович, д.ф.-м.н., професор, затверджені наказом по університету від «02» листопада 2020р. № 3182-с

**2. Термін подання студентом дисертації:** 15 грудня 2020 р.

**3. Об'єкт дослідження:** набір даних щодо купівлі певних продуктів клієнтами

**4. Предмет дослідження:** застосування нейронних мереж у напрямку максимізації вартості клієнтів в сфері роздрібної торгівлі

**5. Перелік завдань, які потрібно розробити:**

1. Огляд предметної області
2. Формування та стратифікація вибірки в рамках гіпотези
3. Розробка нейронної мережі
4. Застосування колаборативних фільтрів
5. Розробка програмного комплексу, що забезпечуватиме використання існуючих та розроблених методів для вирішення задачі

**6. Орієнтовний перелік графічного (ілюстративного) матеріалу:**

- 1). Короткий опис колаборативного фільтру
- 2). Приклади функціонування створеного програмного продукту
- 3). Таблиці у розділі стартап-проекту

**7. Дата видачі завдання:** 10 вересня 2020 року

**Календарний план**

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації
1.	Концептуальний вступ дисертації. Формулювання об'єкта, предмета, цілі, завдань, новизни, практичної значущості результатів	18.09.2020—20.09.2020
2.	Перший розділ. Аналіз предметної області	21.09.2020—30.09.2020
3.	Другий розділ. Розробка нейронної мережі, стратифікації та застосування колаборативного фільтра	01.10.2020—09.11.2020
4.	Третій розділ. Проведення експерименту	10.11.2020—16.11.2020
5.	Четвертий розділ. Стартаппроект	17.11.2020—20.11.2020
6.	Концептуальні висновки. Перспективи розвитку отриманих рішень	21.11.2020—30.11.2020

Студент

В.О. Свириденко

Науковий керівник дисертації

О.С. Макаренко

## РЕФЕРАТ

Магістерська дисертація: 64 с., 14 рис., 18 табл., 1 додаток, 19 джерел.

Тема магістерської дисертації «Система максимізації вартості клієнта (МВК) із використанням нейронних мереж у сфері роздрібно́ї торгівлі».

Актуальність магістерської дисертації обумовлена динамічними змінами у сфері роздрібно́ї торгівлі, що почалися та продовжуються за рахунок появи пандемії COVID-19. Збільшилась потреба в отриманні високої якості аналізу поведінки клієнтів, виявлення тенденцій та використання математичної бази задля підвищення вартості комунікації для клієнта .

Об'єкт дослідження - нейронні мережі, їх можливості та перспективи у сфері роздрібно́ї торгівлі.

Предмет дослідження - моделі та методи застосування нейронних мереж для задачі прогнозування та застосування колаборативних фільтрів для збільшення значення конверсії.

Метою магістерської дисертації є підвищення якісних характеристик роботи нейромереж для розв'язання задач прогнозування поведінки клієнта.

Програмний продукт реалізований за допомогою мови програмування Python 3.7 у середовищі розробки Jupyter Notebook.

Отримані результати: розроблено нейронну мережу та реалізовано застосування колаборативної фільтрації, створені інструменти стратифікації на основі яких проведено аналіз отриманих конверсій.

НЕЙРОННІ МЕРЕЖІ, КОЛАБОРАТИВНИЙ ФІЛЬТР, ПРОГНОЗУВАННЯ, СТРАТИФІКАЦІЯ, КОНВЕРСІЯ.

## ABSTRACT

Master's thesis explanatory note: 64 p., 14 fig., 18 tabl., 1 application, 19 references.

Theme of the master's dissertation " Customer value maximization (CVM) system using neural networks in the field of retail".

The relevance of the master's thesis is due to the dynamic changes in the field of retail, which began and continues due to the pandemic COVID-19. There is an increasing need to obtain high quality analysis of customer behavior, identify trends and use the mathematical basis to increase the cost of communication for the customer.

The object of research is neural networks, their capabilities and prospects in the field of retail.

The subject of research is models and methods of application of neural networks for the problem of forecasting and application of collaborative filters to increase the value of conversion.

The purpose of the master's dissertation is to improve the quality characteristics of neural networks to solve problems of predicting customer behavior.

The software product is implemented using the Python 3.7 programming language in the Jupyter Notebook development environment.

The results are obtained: a neural network is developed and the application of collaborative filtering is implemented, stratification tools are created on the basis of which the analysis of the received conversions is carried out.

NEURAL NETWORKS, COLLABORATIVE FILTER, FORECASTING, STRATIFICATION, CONVERSION.

## ЗМІСТ

ВСТУП .....	8
1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ.....	9
1.1. Огляд літератури.....	9
1.2. Типи CRM.....	9
1.3. Вплив на задоволеність споживачів .....	11
1.4. Максимізація вартості клієнта (МВК).....	14
1.5. Висновки до розділу .....	16
2 ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ. СТРАТИФІКАЦІЯ ТА КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ.....	17
2.1 Нейронні мережі.....	17
2.1.1. Як працює нейронна мережа .....	18
2.1.2. Функція активації .....	19
2.2 Стратифікація даних.....	21
2.1.1. Переваги .....	23
2.1.2. Недоліки .....	24
2.2. Колаборативна фільтрація .....	25
2.3. Висновки до розділу .....	30
3 ОГЛЯД ПРОГРАМНОГО ПРОДУКТУ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ.....	31
3.1. Вхідні данні та навчання моделі .....	31
3.2. Сегментація, формування та стратифікація вибірки .....	34
3.3. Застосування нейронної мережі та колаборативної фільтрації.....	35
3.5. Основні використані бібліотеки .....	38
3.6 Аналіз результатів .....	39
3.7 Висновки до розділу.....	40
4 РОЗРОБКА СТАРТАП-ПРОЕКТУ .....	41
4.1. Опис ідеї проекту. ....	41
4.2. Технологічний аудит проекту. ....	42
4.3. Аналіз ринкових можливостей запуску стартап-проекту. ....	43
4.4. Розроблення ринкової стратегії проекту.....	53
4.5. Розроблення маркетингової програми стартап-проекту .....	56

4.6. Висновки до розділу.....	58
ВИСНОВКИ.....	59
ПЕРЕЛІК ПОСИЛАНЬ .....	60
ДОДАТОК А ЛІСТИНГ ПРОГРАММИ.....	62

## ВСТУП

Під час пандемії COVID-19 клієнтська база у сфері роздрібної торгівлі значно збільшилась та звичайно збільшилась доля не активних клієнтів. Тому очевидно, що першочерговими цілями для кожної компанії буде:

- утримання нових клієнтів;
- повернення старих клієнтів.

Для цього потрібно проаналізувати існуючі підходи в управлінні клієнтських взаємовідносин (CRM) та обміркувати які б нові технології можна було б застосувати для вирішення даних задач. Всі ці данні повинні супроводжуватись математичним апаратом, щоб отриманні результати можна було об'єктивно проаналізувати.

Прогнозування поведінки клієнта не легке завдання, а тим паче спробувати вплинути на його поведінку або звичку. У цьому дослідженні ми спробуємо дослідити ринок на існування інструментів, що допомогли б у вирішенні цієї задачі. Також розробити та імплементувати реалізований алгоритм в рамках нашої задачі. Висунемо декілька гіпотез, які надалі перевіримо та оцінимо їх ефективність за допомогою показника конвертації.



## 1 ОГЛЯД ПРЕДМЕТНОЇ ОБЛАСТІ

### 1.1. Огляд літератури.

Управління відносинами з клієнтами (CRM) - це один із багатьох різних підходів, що дозволяють компанії управляти та аналізувати власну взаємодію зі своїми минулими, поточними та потенційними клієнтами. Він використовує аналіз даних про історію споживачів із компанією для поліпшення ділових відносин із замовниками, зокрема, зосереджуючись на утриманні споживачів і, зрештою, на стимулюванні зростання продажів.

Одним із важливих аспектів підходу до CRM є системи CRM, що збирають дані з різних каналів комунікації, включаючи веб-сайт компанії, телефон, електронну пошту, чат у прямому ефірі, маркетингові матеріали та нещодавно соціальні медіа. Завдяки підходу CRM та системам, що використовуються для його сприяння, підприємства дізнаються більше про свою цільову аудиторію та як найкраще задовольнити їх потреби.[1]

### 1.2. Типи CRM

#### 1. Стратегічний

Стратегічний CRM зосереджений на розвитку бізнес-культури, орієнтованої на споживача.

#### 2. Оперативні

Основною метою систем управління відносинами з клієнтами є інтеграція та автоматизація продажів, маркетингу та підтримки споживачів. Отже, ці системи зазвичай мають інформаційну панель, яка дає загальний огляд трьох функцій в

одному перегляді клієнта, одну сторінку для кожного клієнта, яку може мати компанія. Інформаційна панель може містити інформацію про клієнта, минулі продажі, попередні маркетингові зусилля та багато іншого, узагальнюючи всі відносини між клієнтом та фірмою. Операційний CRM складається з 3 основних компонентів: автоматизація відділу збуту, автоматизація маркетингу та автоматизація послуг:

- а) автоматизація відділу продажів працює на всіх етапах циклу продажів - від початкового введення контактної інформації до перетворення потенційного клієнта в реального клієнта. Він реалізує аналіз стимулювання збуту, автоматизує відстеження історії рахунку клієнта для повторних продажів або майбутніх продажів та координує продажі, маркетинг, кол-центри та торгові точки. Це запобігає дублюванню зусиль між продавцем та клієнтом, а також автоматично відстежує всі контакти та подальші дії між обома сторонами;
- б) автоматизація маркетингу фокусується на полегшенні загального маркетингового процесу, щоб зробити його більш результативним та результативним. Інструменти CRM з можливостями автоматизації маркетингу можуть автоматизувати повторювані завдання, наприклад, розсилати автоматизовані маркетингові електронні листи клієнтам у певний час або розміщувати маркетингову інформацію в соціальних мережах. Мета автоматизації маркетингу - перетворити лідера продажів на повноцінного клієнта. Системи CRM сьогодні також працюють над залученням клієнтів через соціальні мережі;
- в) автоматизація послуг - це частина CRM-системи, яка орієнтована на безпосередню технологію обслуговування споживачів. Завдяки автоматизації послуг клієнти отримують підтримку за допомогою багатьох каналів, таких як телефон, електронна пошта, бази знань, портали продажу квитків, поширені запитання та багато іншого.

### 3. Аналітичний

Роль аналітичних CRM-систем полягає в аналізі даних про клієнтів, зібраних через безліч джерел, і представлення їх для того, щоб менеджери підприємств могли приймати більш обґрунтовані рішення. Аналітичні CRM-системи використовують такі методи, як аналіз даних, кореляція та розпізнавання шаблонів для аналізу даних клієнта. Ця аналітика допомагає покращити обслуговування клієнтів, знаходячи невеликі проблеми, які можна вирішити, можливо, по-різному збуту для різних частин споживчої аудиторії. Наприклад, шляхом аналізу купівельної поведінки клієнтської бази компанія може побачити, що ця клієнтська база останнім часом не купує багато продуктів. Просканувавши ці дані, компанія може подумати продавати для цієї групи споживачів інакше, щоб найкраще повідомити, як продукція цієї компанії може бути корисною для цієї групи.

### 4. Спільна

Третьою основною метою CRM-систем є залучення зовнішніх зацікавлених сторін, таких як постачальники, постачальники та дистриб'ютори, та обмін інформацією про клієнтів між групами / відділами та організаціями. Наприклад, зворотній зв'язок може бути зібраний за допомогою дзвінків технічної підтримки, що може допомогти надавати вказівки щодо збуту продуктів та послуг конкретному клієнту в майбутньому.[2,3]

### 1.3. Вплив на задоволеність споживачів

За допомогою систем управління відносинами з клієнтами клієнти отримують кращі послуги у повсякденному процесі. З більш надійною інформацією їх попит на самообслуговування з боку компаній зменшиться. Якщо менше необхідності

взаємодіяти з компанією для вирішення різних проблем, рівень задоволеності споживачів зростає. Ці основні переваги CRM будуть гіпотетично пов'язані з трьома видами власного капіталу, які є відносинами, вартістю та торговою маркою, і, зрештою, з капіталом клієнта. Вісім переваг було визнано для забезпечення рушіїв вартості:

1. Розширена здатність орієнтуватися на вигідних клієнтів.
2. Інтегрована допомога через різні канали.
3. Підвищена ефективність та ефективність продажів.
4. Поліпшення цін.
5. Індивідуальні товари та послуги.
6. Покращена ефективність та результативність обслуговування клієнтів.
7. Індивідуалізовані маркетингові повідомлення також називають кампаніями.
8. Поєднуйте клієнтів та всі канали на одній платформі.

У 2012 році, переглянувши попередні дослідження, хтось обрав деякі з тих переваг, які є більш важливими для задоволення споживачів, та узагальнив їх у такі випадки:

1. Покращення обслуговування клієнтів: загалом, у клієнтів виникають певні запитання, зауваження чи запити. Послуги CRM надають компанії змогу виробляти, розподіляти та управляти запитами або чимось, що надходить від клієнтів. Наприклад, програмне забезпечення кол-центру, яке допомагає зв'язати клієнта з менеджером або особою, яка може найкраще допомогти їм у існуючій проблемі, є однією з можливостей CRM, яку можна застосувати для підвищення ефективності.
2. Покращений персоніфікований сервіс або індивідуальний сервіс: Персоналізація обслуговування клієнтів або індивідуальний сервіс надає компаніям можливість покращити розуміння та отримання знань клієнтів, а також краще знати про уподобання, вимоги та вимоги своїх клієнтів. (CVM)
3. Відповідаючи потребам замовника: ситуації та потреби споживачів можуть бути зрозумілі фірмами, орієнтуючись на потреби та вимоги замовника.

4. Сегментація клієнтів: У CRM сегментація використовується для класифікації клієнтів відповідно до певної подібності, наприклад, галузі, роботи чи деяких інших характеристик, на подібні групи. Хоча ці характеристики можуть бути одним або кількома атрибутами. Це може бути визначено як підрозділ клієнтів на основі вже відомого хорошого дискримінатора.
5. Поліпшення кастомізації маркетингу. Сенс кастомізації маркетингу полягає в тому, що фірма чи організація адаптуються та змінюють свої послуги чи товари на основі представлення різних та унікальних продуктів чи послуг для кожного клієнта. З метою забезпечення задоволення потреб і вимог замовника організація використовує налаштування. Компанії можуть інвестувати в інформацію від клієнтів, а потім налаштовувати свої товари чи послуги для підтримки інтересів клієнтів.
6. Багатоканальна інтеграція: багатоканальна інтеграція показує суть спільного створення цінності споживача в CRM. З іншого боку, вміння компанії успішно виконувати багатоканальну інтеграцію сильно залежить від здатності організації збирати інформацію про клієнтів з усіх каналів та включати її з іншою відповідною інформацією.
7. Заощадження часу: CRM дозволить компаніям частіше взаємодіяти з клієнтами за допомогою персоналізованого способу обміну повідомленнями та комунікаціями, який можна швидко виробляти та своєчасно узгоджувати, і нарешті вони зможуть краще зрозуміти своїх клієнтів і, отже, сподіватися на їх потреби.
8. Поліпшення знань клієнтів: фірми можуть створювати та вдосконалювати продукти та послуги за допомогою інформації від відстеження (наприклад, за допомогою відстеження веб-сайтів) поведінки клієнтів до смаків та потреб клієнтів. CRM може сприяти досягненню конкурентних переваг у покращенні здатності фірми збирати інформацію про клієнтів для налаштування продуктів та послуг відповідно до потреб клієнтів.

#### 1.4. Максимізація вартості клієнта (МВК)

Максимізація споживчої вартості - це набір методів та дій, що використовуються для спонукання споживачів збільшити частоту та обсяг їх транзакцій та збільшити тривалість часу, коли вони залишаються активними клієнтами бізнесу.

Після того, як компанія придбала нового клієнта, кінцевою метою зусиль з маркетингу та утримання є максимізація доходів, які клієнт створює для компанії (припускаючи стабільну залежність між доходом та прибутком). Три основні фактори, що сприяють загальному доходу, який генерує будь-який конкретний клієнт, - це час (скільки часу клієнт залишається активним клієнтом), частота покупок (як часто клієнт купує щось у компанії) та грошова вартість покупок (скільки гроші, які клієнт витрачає з компанією). Таким чином, максимізація вартості клієнта для бізнесу означає максимізацію часу  $\times$  частоти покупок  $\times$  рівняння грошової вартості.

##### 1.4.1 Як досягти максимізації вартості споживача.

Надання споживачам продуктів чи послуг, які вони шукають, за конкурентоспроможними цінами та надзвичайним обслуговуванням споживачів, очевидно, є найкращим способом забезпечити, щоб замовник залишався в компанії на довгий термін і продовжував купувати у неї. Однак компанія завжди має певні дії, які заохочуватимуть клієнтів триматися довше і витратити більше. Проблема полягає в тому, щоб знати, які дії застосовувати до яких клієнтів і коли це робити для досягнення максимальних результатів.

1. Сегментація клієнтів - важливо сегментувати клієнтів на невеликі групи та звертатися до окремих клієнтів на основі фактичної поведінки - замість жорсткого кодування будь-яких заздалегідь продуманих понять, роблячи припущення про те, що робить клієнтів схожими один на одного або дивлячись лише на зведені / усереднені дані (що приховує важливі факти про окремих клієнтів).
2. Відстеження клієнтів з часом - Дуже важливо стежити за тим, як клієнти переміщуються між різними сегментами з часом (тобто динамічною сегментацією), включаючи контекст життєвого циклу клієнта та аналіз когорт - замість того, щоб просто визначати, в яких сегментах зараз перебувають клієнти, не враховуючи, як і коли вони прибув туди.
3. Точне прогнозування поведінки майбутніх клієнтів (наприклад, конвертувати, відбивати, витратити більше, витратити менше) - Завжди слід використовувати методи прогнозування поведінки клієнтів - замість того, щоб просто дивитись у дзеркало заднього виду історичних даних.
4. Час “життя” клієнта (LTV) - Моделювання максимізації вартості клієнта повинно базуватися на використанні розширених розрахунків для визначення вартості життя клієнта (LTV) кожного клієнта та на основі прийняття рішень на ньому - замість того, щоб дивитись лише на короткостроковий дохід, який клієнт може привести компанію.
5. Оптимізація маркетингових дій - маркетолог або експерт з утримання повинен на основі об’єктивних метрик знати, які саме маркетингові дії робити зараз для кожного клієнта, щоб максимізувати довгострокову цінність кожного клієнта - замість того, щоб намагатися з’ясувати, що робити на основі інформаційної панелі або купи звітів.

### 1.5. Висновки до розділу

Останнім часом, з метою збільшення прибутків компаній набирають популярність різні підходи в управлінні відносинами з клієнтами (CRM). Досягання цієї мети може мати безліч різних підходів, та все ж ми сконцентруємось на напрямку максимізації вартості клієнта (CVM).

Для досягнення цієї мети ми сконцентруємось на наступному:

- а) сегментація клієнтів;
- б) висунення гіпотез;
- в) аналіз та збільшення значення LTV (lifetime value), час життя клієнта;
- г) аналіз маркетингових дій та їх подальше покращення та оптимізація.



## 2 ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ. СТРАТИФІКАЦІЯ ТА КОЛАБОРАТИВНА ФІЛЬТРАЦІЯ

### 2.1 Нейронні мережі

Нейронна мережа - це мережа або ланцюг нейронів, або в сучасному розумінні штучна нейронна мережа, що складається із штучних нейронів або вузлів. Таким чином, нейронна мережа - це або біологічна нейронна мережа, що складається з реальних біологічних нейронів, або штучна нейронна мережа для вирішення проблем штучного інтелекту (ШІ). Зв'язки біологічного нейрона моделюються як ваги. Позитивна вага відображає збудливий зв'язок, тоді як негативні значення означають гальмівні зв'язки. Усі входи модифікуються вагою та підсумовуються. Ця діяльність називається лінійною комбінацією. Нарешті, функція активації контролює амплітуду вихідного сигналу. Наприклад, прийнятний діапазон виходу зазвичай становить від 0 до 1, а може бути від -1 до 1.

Ці штучні мережі можуть бути використані для прогнозного моделювання, адаптивного управління та додатків, де їх можна навчити за допомогою набору даних. Самонавчання, спричинене досвідом, може відбуватися всередині мереж, що може робити висновки із складної і, здавалося б, не пов'язаної набору інформації. На рисунку 2.1 зображено приклад ймовірнісної нейронної мережі.[4]

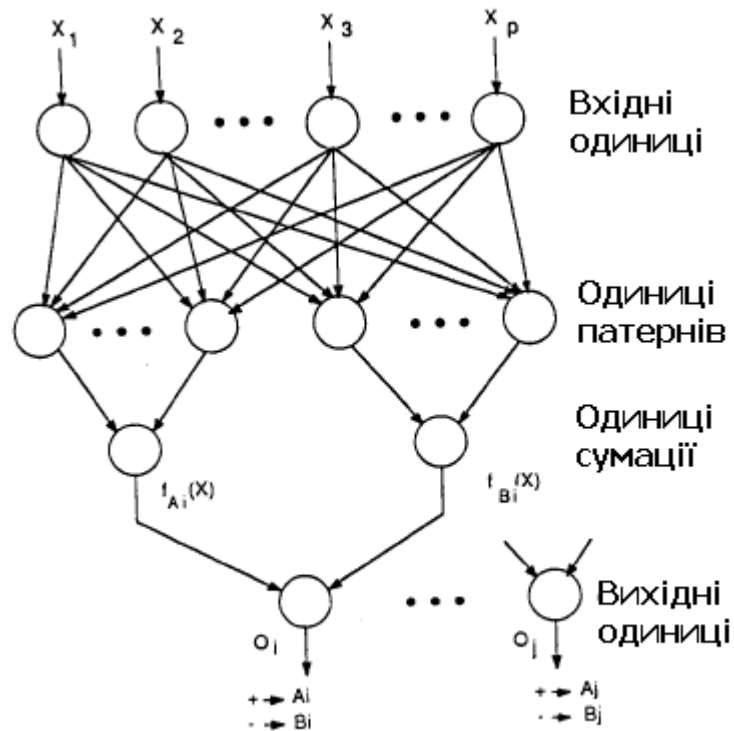
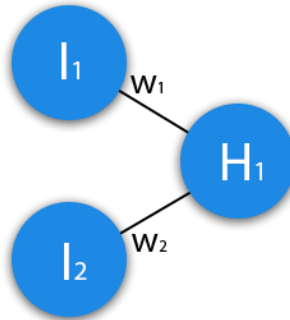


Рисунок. 2.1 - Ймовірнісна нейронна мережа

### 2.1.1. Як працює нейронна мережа

На рисунку 2.2 зображена частина нейронної мережі, де буквами  $I$  позначені вхідні нейрони, буквою  $H$  - прихований нейрон, а буквою  $w$  - ваги. З формули видно, що вхідна інформація - це сума всіх вхідних даних, помножених на відповідні їм ваги. Тоді дамо на вхід 1 і 0. Нехай  $w_1 = 0.4$  і  $w_2 = 0.7$ . Вхідні дані нейрона  $H_1$  будуть наступними:  $1 * 0.4 + 0 * 0.7 = 0.4$ . Отже, коли у нас є вхідні дані, ми можемо отримати вихідні дані, підставивши вхідні значення в функцію активації (докладніше про неї далі). Тепер, коли у нас є вихідні дані, ми передаємо їх далі. І так, ми повторюємо для всіх верств, поки не дійдемо до вихідного нейрона. Запустивши таку мережу в перший раз ми побачимо, що відповідь далекий від

правильно, тому що мережа не натренована. Щоб поліпшити результати ми будемо її тренувати. Але перш ніж дізнатися як це робити, давайте введемо кілька термінів і властивостей нейронної мережі.



$$1) H_{1input} = (I_1 * w_1) + (I_2 * w_2)$$

$$2) H_{1output} = f_{activation}(H_{1input})$$

Рисунок 2.2 – Вхідні і прихований нейрони

### 2.1.2. Функція активації

Функція активації - це спосіб нормалізації вхідних даних (ми вже говорили про це раніше). Тобто, якщо на вході у вас буде велика кількість, пропустивши його через функцію активації, ви отримаєте вихід в потрібному вам діапазоні. Функцій активації досить багато тому ми розглянемо найосновніші: Лінійна, сигмоїда (Логістична) і Гіперболічний тангенс. Головні їх відмінності - це діапазон значень.

Функція (Рис. 2.3) майже ніколи не використовується, за винятком випадків, коли потрібно протестувати нейронну мережу або передати значення без перетворень.

$$f(x) = x$$

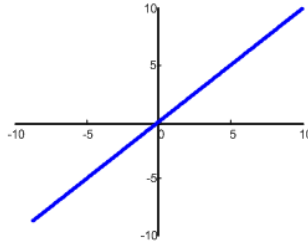


Рисунок 2.3 – Лінійна функція

Функція (Рис. 2.4) найпоширеніша функція активації, її діапазон значень  $[0,1]$ . Саме на ній показано більшість прикладів в мережі, також її іноді називають логістичною функцією. Відповідно, якщо в вашому випадку присутні негативні значення (наприклад, акції можуть йти не тільки вгору, але і вниз), то вам знадобиться функція яка захоплює і негативні значення.

$$f(x) = \frac{1}{1+e^{-x}}$$

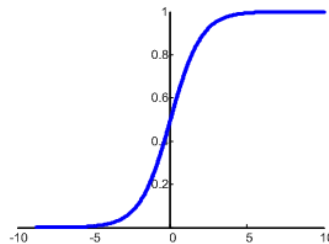


Рисунок 2.4 – Сигмоїда

Має сенс використовувати гіперболічний тангенс (Рис. 2.5), тільки тоді, коли ваші значення можуть бути і негативними, і позитивними, так як діапазон функції  $[-1,1]$ . Використовувати цю функцію тільки з позитивними значеннями недоцільно так як це значно погіршить результати вашої нейромережі.

$$f(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

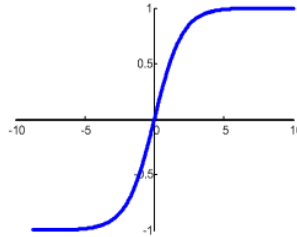


Рисунок 2.5 – Гіперболічний тангенс

## 2.2 Стратифікація даних

У статистиці стратифікована вибірка - це метод вибірки з сукупності, яку можна розділити на субпопуляції.

У статистичних дослідженнях, коли субпопуляції в межах загальної сукупності різняться, може бути вигідним відбирати кожну субпопуляцію (прошарок) незалежно. Стратифікація - це процес поділу представників популяції на однорідні підгрупи перед відбором проб. Пласти повинні визначати поділ населення. Тобто вона повинна бути колективно вичерпною та взаємовиключною: кожен елемент у сукупності повинен бути віднесений до однієї і тієї самої верстви. Потім у межах кожного шару застосовується проста довільна вибірка або систематична вибірка. Метою є підвищення точності вибірки за рахунок зменшення помилки вибірки. Він може отримати середньозважене значення, яке має меншу мінливість, ніж середнє арифметичне простої випадкової вибірки сукупності.

У обчислювальній статистиці стратифікована вибірка є методом зменшення дисперсії, коли методи Монте-Карло використовуються для оцінки статистики популяції з відомої сукупності.

Припустимо, що нам потрібно оцінити середню кількість голосів за кожного кандидата на виборах. Припустимо, що в країні 3 міста: у місті А - 1 мільйон робітників заводу, у місті Б - 2 мільйони офісних працівників, а у місті С - 3 мільйони пенсіонерів. Ми можемо обрати випадкову вибірку розміром 60 для всієї сукупності, але існує певна ймовірність того, що отримана випадкова вибірка погано збалансована в цих містах і, отже, є упередженою, що спричиняє значну помилку в оцінці. Натомість, якщо ми вирішимо взяти випадкову вибірку 10, 20 та 30 з міст А, В та С відповідно, тоді ми можемо допустити меншу помилку в оцінці для того самого загального обсягу вибірки. Цей метод зазвичай використовується, коли популяція не є однорідною групою.

Середнє значення та дисперсія стратифікованої випадкової вибірки наведені як:

$$\bar{x} = \frac{1}{N} \sum_{h=1}^L N_h x_h$$

$$s_x^2 = \sum_{h=1}^L \left(\frac{N_h}{N}\right)^2 \left(\frac{N_h - n_h}{N_h}\right) \frac{s_h^2}{n_h}$$

де  $L$  - кількість шарів

$N$  - сума всіх розмірів шару

$N_h$  - розмір шару  $h$

$x_h$  - зразок середнього шару  $h$

$n_h$  - кількість спостережень у прошарку  $h$

$s_h$  - зразок стандартного відхилення прошарку  $h$

### 2.1.1. Переваги

Причини використання стратифікованої вибірки замість простої випадкової вибірки включають:

1. Якщо вимірювання всередині шарів мають нижче стандартне відхилення, стратифікація дає меншу похибку в оцінці.
2. Для багатьох застосувань вимірювання стають більш керованими та / або дешевшими, коли популяція групується у шари.
3. Часто бажано мати оцінки параметрів популяції для груп всередині популяції.

Якщо щільність населення сильно варіюється в межах регіону, стратифікована вибірка забезпечить можливість проведення оцінок з однаковою точністю в різних частинах регіону та порівняння субрегіонів з однаковою статистичною потужністю. Наприклад, в Онтаріо під час опитування, проведеного по всій провінції, може бути використана більша частка вибірки на менш населеній півночі, оскільки диспропорція в кількості населення між північню та півднем настільки велика, що частка вибірки на основі провінційної вибірки в цілому може призвести до збір лише небагатьох даних із півночі.

### 2.1.2. Недоліки

Стратифікована вибірка не є корисною, коли сукупність не може бути вичерпно розподілена на розділені підгрупи. Було б неправильним застосуванням методики зробити розмір вибірки підгруп пропорційним обсягу даних, доступних з підгруп, а не масштабувати розміри вибірки до розмірів підгруп (або їх дисперсій, якщо відомо, що вони суттєво відрізняються - наприклад, за допомогою F Тест). Дані, що представляють кожну підгрупу, приймаються однаково важливими, якщо підозра на їх зміну вимагає стратифікованої вибірки. Якщо дисперсії підгрупи суттєво різняться, і дані потрібно розкласти на дисперсію, неможливо одночасно зробити розмір вибірки кожної підгрупи пропорційним розміру підгрупи в межах загальної сукупності. Ефективний спосіб розподілу ресурсів вибірки між групами, які різняться за своїми засобами, дисперсією та витратами, див. У розділі "Оптимальне розподіл". Проблема стратифікованої вибірки у випадку невідомих пріоритетів класу (співвідношення субпопуляцій у всій сукупності) може мати шкідливий вплив на результати будь-якого аналізу набору даних, наприклад класифікація. У зв'язку з цим коефіцієнт вибірки мінімакс може бути використаний, щоб зробити набір даних надійним щодо невизначеності основного процесу генерації даних.

Поєднання підшарів для забезпечення адекватної кількості може призвести до парадоксу Сімпсона, коли тенденції, які насправді існують у різних групах даних, зникають або навіть зворотні при об'єднанні груп.



## 2.2. Колаборативна фільтрація

Колаборативна фільтрація (CF) - це техніка, що використовується системами рекомендацій. Колаборативна фільтрація має два сенсу, вузький і більш загальний.

У новому, вузькому розумінні, колаборативна фільтрація - це метод автоматичного прогнозування (фільтрації) щодо інтересів користувача шляхом збору переваг або інформації про смак від багатьох користувачів (співпраця). Основним припущенням підходу до спільної фільтрації є те, що якщо особа А дотримується такої ж думки, як і людина Б щодо того чи іншого питання, то А, швидше за все, має думку В щодо іншого питання, ніж думка випадково обраної людини. Наприклад, система спільної фільтрації рекомендацій щодо телевізійних смаків може робити прогнози щодо того, яке телевізійне шоу сподобається користувачеві, даючи частковий перелік смаків цього користувача (симпатії чи антипатії). Зверніть увагу, що ці передбачення стосуються лише користувача, але використовують інформацію, отриману багатьма користувачами. Це відрізняється від простішого підходу надання середнього (неспецифічного) бала для кожного предмета, що цікавить, наприклад, на основі його кількості голосів.

У більш загальному розумінні спільна фільтрація - це процес фільтрації інформації або шаблонів із використанням методів, що передбачають співпрацю між кількома агентами, точками зору, джерелами даних тощо. Застосування спільної фільтрації зазвичай включає дуже великі набори даних. Методи спільної фільтрації застосовувались до багатьох різних типів даних, включаючи: дані зондування та моніторингу, такі як при дослідженні корисних копалин, зондування навколишнього середовища на великих площах або декілька датчиків; фінансові дані, такі як установи фінансових послуг, які інтегрують багато фінансових джерел; або в електронній торгівлі та веб-додатках, де основна увага приділяється даним

користувачів тощо. Решта цього обговорення зосереджується на спільній фільтрації даних користувачів, хоча деякі методи та підходи можуть застосовуватися і до інших основних додатків. (Рис. 2.6)

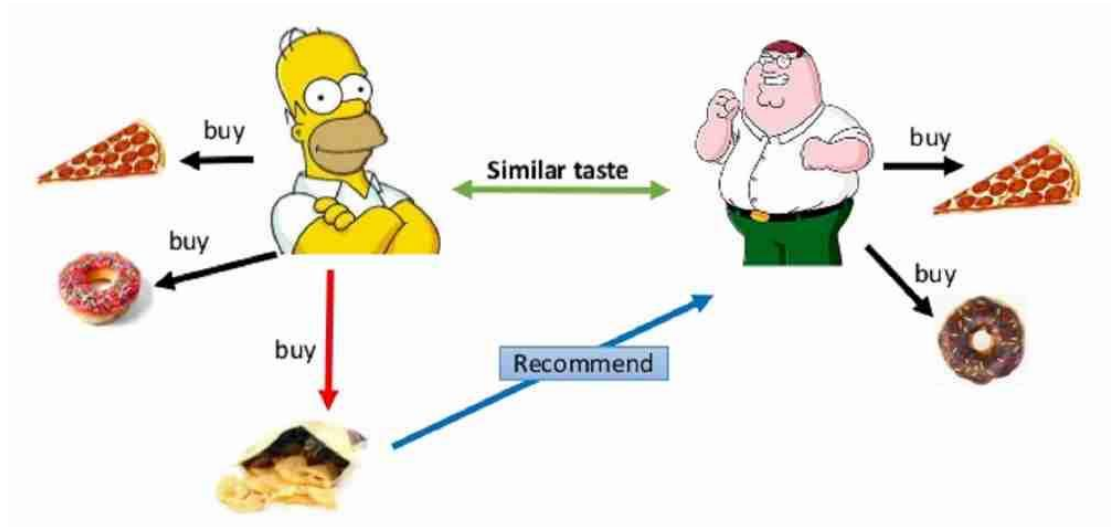


Рисунок 2.6 – Приклад колаборативної фільтрації

Колаборативні системи фільтрації мають багато форм, але багато загальних систем можна звести до двох етапів:

- 1) шукайте користувачів, які поділяють однакові шаблони рейтингу з активним користувачем (користувачем, для якого передбачено прогнозування);
- 2) використовуйте оцінки користувачів-однодумців, знайдені на кроці 1, для розрахунку прогнозу для активного користувача.

Це підпадає під категорію колаборативної фільтрації на основі користувачів. Конкретним застосуванням цього є користувальницький алгоритм "Найближчий сусід".

Крім того, колаборативна фільтрація на основі товарів (користувачі, які придбали  $x$ , також купили  $y$ ), здійснюється орієнтовано на товар:

1. Побудуйте матрицю елемент-елемент, що визначає взаємозв'язки між парами елементів.

2. Визначте смаки поточного користувача, вивчивши матрицю та зіставивши дані цього користувача.

Інша форма колаборативної фільтрації може базуватися на неявних спостереженнях за нормальною поведінкою користувача (на відміну від штучної поведінки, накладеної рейтинговим завданням). Ці системи спостерігають за тим, що робив користувач разом із тим, що робили всі користувачі (яку музику вони слухали, які предмети вони придбали), і використовують ці дані для прогнозування поведінки користувача в майбутньому або для прогнозування того, як може сподобатися користувачеві поводитися, даючи можливість. Потім ці прогнози потрібно відфільтрувати за допомогою бізнес-логіки, щоб визначити, як вони можуть вплинути на дії бізнес-системи. Наприклад, не корисно пропонувати комусь продати певний музичний альбом, якщо він уже продемонстрував, що володіє цією музикою.

Спираючись на систему оцінювання або рейтинг, яка є усередненою для всіх користувачів, ігнорує конкретні вимоги користувача і особливо погана у завданнях, де є великі розбіжності в інтересах (як у музичній рекомендації). Однак існують інші методи боротьби з вибухом інформації, такі як веб-пошук та кластеризація даних.

Підхід, заснований на пам'яті, використовує дані рейтингу користувачів для обчислення подібності між користувачами або елементами. Типовими прикладами цього підходу є CF на основі сусідства та рекомендації top-N на основі предметів / користувачів. Наприклад, у підходах, заснованих на користувачах, значення оцінок, які користувач  $u$  дає елементу  $i$ , обчислюється як сукупність рейтингу подібних користувачів щодо елемента:

$$r_{u,i} = \text{aggr}_{w \in U} r_{w,i}$$

де  $U$  позначає набір найпопулярніших  $N$  користувачів, які найбільш схожі на користувача  $u$ , який оцінив елемент  $i$ . Деякі приклади функції агрегування включають:

$$r_{u,i} = \frac{1}{N} \sum_{w \in U} r_{w,i}$$

$$r_{u,i} = k \sum_{w \in U} \text{simil}(u, w) r_{w,i}$$

де  $k$  - коефіцієнт нормування, визначений як:

$$k = \frac{1}{\sum_{w \in U} |\text{simil}(u, w)|}$$

$$r_{u,i} = r_u + k \sum_{w \in U} \text{simil}(u, w) (r_{w,i} + r_w)$$

де  $r_w$  - середня оцінка користувача  $u$  для всіх елементів, оцінених  $u$ .

Алгоритм, що базується на сусідстві, обчислює схожість між двома користувачами або елементами та виробляє прогноз для користувача, беручи середньозважене значення всіх оцінок. Обчислення подібності між елементами або користувачами є важливою частиною цього підходу. Для цього використовуються різні виміри, такі як кореляція Пірсона та подібність векторних косинусів.

Подібність кореляції Пірсона двох користувачів  $x$ ,  $y$  визначається як:

$$\text{simil}(x, y) = \frac{\sum_{i \in I_{xy}} (r_{x,i} + r_x)(r_{y,i} + r_y)}{\sqrt{\sum_{i \in I_{xy}} (r_{x,i} + r_x)^2} \sqrt{\sum_{i \in I_{xy}} (r_{y,i} + r_y)^2}}$$

де  $I_{xy}$  - набір елементів, оцінений як користувачем  $x$ , так і користувачем  $y$ .

Підхід, заснований на косинусах, визначає подібність косинусів між двома користувачами  $x$  та  $y$  як:

$$simil(x, y) = \cos(x, y) = \frac{x \times y}{||x|| \times ||y||} = \frac{\sum_{i \in I_{xy}} r_{x,i} r_{y,i}}{\sqrt{\sum_{i \in I_x} r_{x,i}^2} \sqrt{\sum_{i \in I_y} r_{y,i}^2}}$$

Алгоритм рекомендацій top-N на основі користувачів використовує векторну модель на основі подібності для ідентифікації  $k$  найбільш подібних користувачів до активного користувача. Після того, як буде знайдено  $k$  найбільш подібних користувачів, їх відповідні матриці елементів користувача об'єднуються для ідентифікації набору елементів, що рекомендуються. Популярним методом пошуку подібних користувачів є хешування, що враховує місцевість, що реалізує механізм найближчого сусіда за лінійний час.

Переваги такого підходу включають: пояснюваність результатів, що є важливим аспектом систем рекомендацій; простота створення та використання; легке полегшення нових даних; незалежність від змісту елементів, що рекомендуються; гарне масштабування з елементами зі спільною оцінкою.

У цього підходу також є кілька недоліків. Його ефективність знижується, коли дані стають рідкими, що часто трапляється з елементами, пов'язаними з Інтернетом. Це заважає масштабованості цього підходу і створює проблеми з великими наборами даних. Хоча він може ефективно обробляти нових користувачів, оскільки він покладається на структуру даних, додавання нових елементів стає більш складним, оскільки це подання, як правило, покладається на конкретний векторний простір. Додавання нових елементів вимагає включення нового елемента та повторного вставлення всіх елементів у структуру.[5]

### 2.3. Висновки до розділу

У вирішенні наших поставлених задач, будуть застосовані наступні інструменти.

1. Проведення кампаній на стратифікованих пілотних та контрольних групах.  
Що були сегментовані за рядом критеріїв.
  - а) розмір чека;
  - б) частота відвідування;
  - в) ціль відвідування.
2. Висунення та перевірка гіпотез, оцінка їх ефективності за допомогою конверсії оціненої на інкрементальному заробітку.
3. Застосування колаборативних фільтрів та нейронної мережі з метою підвищення показників конверсії (більш влучним застосуванням наших пропозицій).

## 3 ОГЛЯД ПРОГРАМНОГО ПРОДУКТУ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

### 3.1. Вхідні данні та навчання моделі

Для навчання нашої нейронної мережі ми використовуємо наступні вхідні данні:

- а) memberID;
- б) кількість чеків;
- в) кількість чеків, що містять певний productid;
- г) productid.

ProductID – ID для кожної унікальної пари name та brand.

Name – певний рівень ієрархічності кожного продукту, наприклад, рівень 1,2.

1) продовольчі товари:

- а) консерви та бакалія;
- б) напої;

2) непродовольчі товари:

- а) хімічні засоби для добу;
- б) нехімічні засоби.

В результаті ми отримаємо певний score, нормалізований від 0 до 3. (табл. 3.1)

Таблиця 3.1 Результат роботи нейронної мережі

memberID	Score	ProductID
753304	2,26	11746
446183	0,56	28353
658967	2,42	84482
140410	1,94	5299

Продовження таблиці 3.1

memberID	Score	ProductID
76996	0,91	84482
859787	2,15	84482
626328	1,02	5299
475568	2,2	5299
548945	2,36	28353
615920	1,64	84482
244311	1,62	84482
63393	0,07	5299
425053	0,54	28353
666767	0,43	11746
896268	2,58	11746
66558	0,98	5299
513684	2,3	28353
287961	2,91	84482
70346	2,83	11746
351294	0,54	28353
172090	0,93	28353
324012	0,38	28353
895745	0,93	28353
999378	1,89	28353
253154	2,48	84482
463805	2,79	5299
393675	1,83	28353
816757	2,04	11746
247248	0,66	11746



Продовження таблиці 3.1

memberID	Score	ProductID
461825	1,05	84482
59639	2,61	11746
482494	2,49	11746
141590	1,3	5299
115386	2,81	84482
965162	0,08	11746
186900	1,71	5299
789303	2,18	11746
542241	0,08	11746
830954	2,71	84482
437556	1,55	28353
443309	2,63	11746
478203	0,74	28353
485968	0,11	11746
400805	0,39	28353
704811	0,03	28353
769126	1,12	84482
214018	1,24	11746
538931	0,34	28353
955653	2,6	28353

### 3.2. Сегментація, формування та стратифікація вибірки

Для проведення нашого експерименту було відібрано 50 тисяч клієнтів певного сегменту та проведено стратифікацію на пілотну (номер 2) та на контрольну (номер 1) групу за наступними критеріями:

- а) середній чек;
- б) кількість чеків.

Отримали незначну розбіжність, яка свідчить про статистично коректну стратифікацію (Рис. 3.1) :

```
campaign.groupby(['audience_id', 'group_id', ])[['avg_cheque']].mean()
```

		avg_cheque
audience_id	group_id	
temp	1	256.347106
	2	256.394342

```
campaign.groupby(['audience_id', 'group_id', ])[['buys_count']].mean()
```

		buys_count
audience_id	group_id	
temp	1	2.070826
	2	2.069466

Рисунок 3.1 – результати стратифікації

### 3.3. Застосування нейронної мережі та колаборативної фільтрації

З метою максимізації вартості запропонованої комунікації клієнту, ми повинні збільшувати відсоток конверсії, тобто відсоток клієнтів, що скористались нашою пропозицією, для цього застосуємо вже навчену нами нейронну мережу.

Відразу отримуємо наступні значення:

Known members : 37\_988

Unknown members : 12\_012

Це свідчить про те, що понад 12 тисяч клієнтів для нас нові, або ми замало знаємо про них. Продовжуємо роботу лише з знайомими для нас. Переходимо до етапу колаборативної фільтрації. Розглянемо приклад з 4 продуктами та 4 клієнтами (Рис. 3.2) :

Клієнти				
	K1	K2	K3	K4
A	😊		😊	😊
B	😊	😊		😞
C	😞	😞		
D		😊	😊	😊

Продукти

Рисунок 3.2 – Приклад дії колаборативної фільтрації, етап 1

Ми бачимо схожість між клієнтами K1 та K2, тому за методом колаборативної фільтрації ми отримаємо рекомендації зображені червоним колом (Рис. 3.3) :

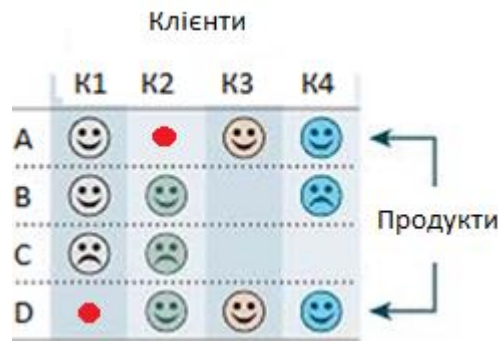


Рисунок 3.3 – Приклад дії колаборативної фільтрації, етап 2

Припустимо, що обидва клієнти залишились задоволеними, тоді наступна ітерація буде виглядати наступним чином (Рис. 3.4):



Рисунок 3.4 – Приклад дії колаборативної фільтрації, етап 3

В результаті застосування колаборативної фільтрації, ми отримали вибірку розміром 4 тисячі, що в подальшому також буде стратифіковано (Рис. 3.5).

```
plt.plot(range(0,len(df_member_productid)),df_member_productid.score)
[<matplotlib.lines.Line2D at 0x2958cf5e788>]
```

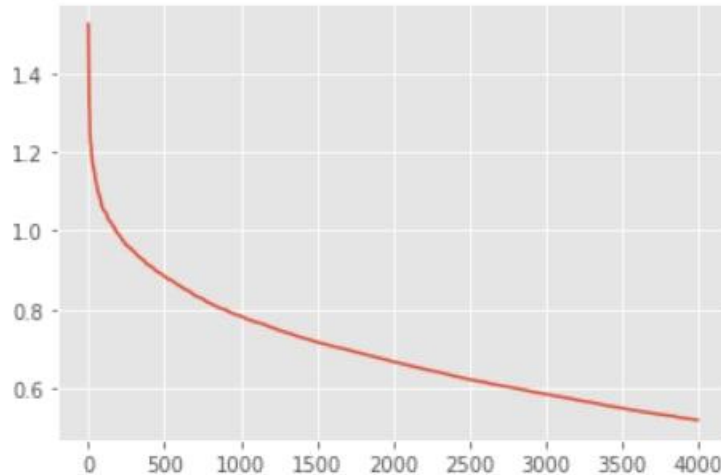


Рисунок 3.5 – Розподіл значень score

### 3.4. Вибір мови програмування

Мова програмування обиралась за наступними критеріями:

- а) сучасність;
- б) можливість розроблювати і запускати застосування у різних операційних;
- в) системах (кросплатформеність);
- г) наявність бібліотек для класифікації, аналізу та візуалізації даних;
- д) можливість легко взаємодіяти з БД SQL.

Згідно з перерахованими критеріями було обрано мову програмування Python, а саме його веб-оболонка Jupyter

### 3.5. Основні використані бібліотеки

Модуль `sys` надає програмісту набір функцій, які дають інформацію про те, як інтерпретатор Python виконує з операційною системою. Модуль `sys` дає наступну інформацію:

- яка версія Пітона запущена;
- шлях до інтерпретатора Python, виконуючий поточний скрипт;
- параметри командного рядка, використовуються при запуску на виконання скрипта;
- прапори, встановленим інтерпретатором;
- подання значень з плаваючою точкою;
- багато іншого.

Pandas - програмна бібліотека мовою Python для обробки і аналізу даних. Робота pandas з даними будується поверх бібліотеки NumPy, що є інструментом нижчого рівня. Надає спеціальні структури даних і операції для маніпулювання числовими таблицями і тимчасовими рядами. Назва бібліотеки походить від економетричного терміна «панельні дані», використовуваного для опису багатовимірних структурованих наборів інформації.

SQLAlchemy - це програмна бібліотека мовою Python для роботи з реляційними СУБД з застосуванням технології ORM. Служить для синхронізації об'єктів Python і записів реляційної бази даних. SQLAlchemy дозволяє описувати структури баз даних і способи взаємодії з ними на мові Python без використання SQL.

Matplotlib - бібліотека на мові програмування Python для візуалізації даних двовимірної (2D) графікою (3D графіка також підтримується). Одержувані зображення можуть бути використані в якості ілюстрацій в публікаціях.

### 3.6 Аналіз результатів

Порівняємо результати сформованих у нас вибірок. На рисунку 3.6 зображено вибірку сформовано без застосування нейронної мережі.

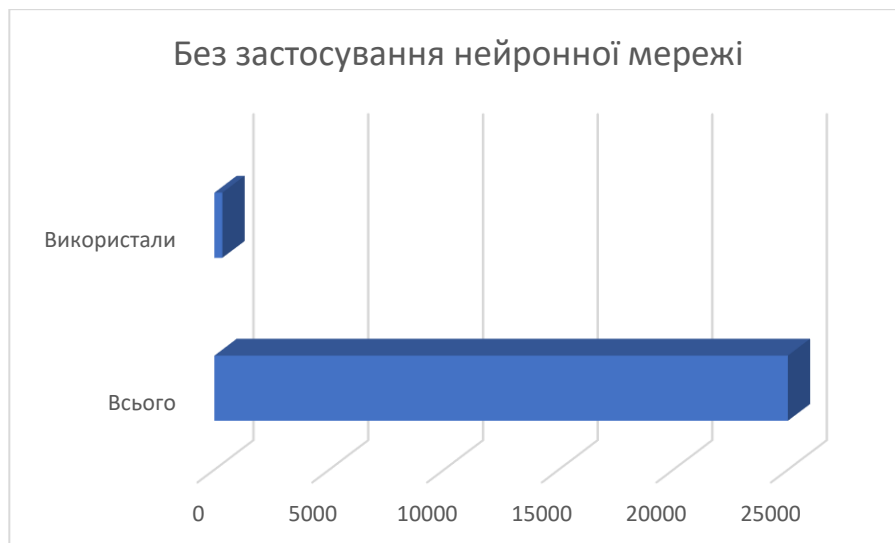


Рисунок 3.6 – Розмір вибірки без застосування нейронної мережі

На рисунок 3.7 зображено вибірку сформовану з застосування нейронної мережі.

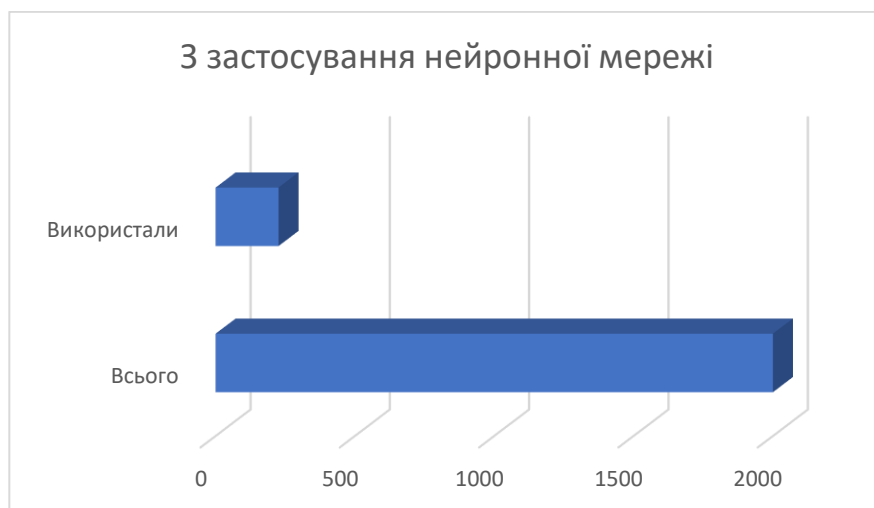


Рисунок 3.7 – Розмір вибірки після застосування нейронної мережі

Для кращого розуміння виведемо показники конверсії для першого та другого випадку (Рисунок 3.8).



Рисунок 3.8 – Конверсія за результатами перевірки гіпотези

Отримані конверсії відрізняються в понад 8 разів. Набагато краще себе показав алгоритм з застосування нейронної мережі.

### 3.7 Висновки до розділу.

Отже, була реалізована програма на мові програмування Python в веб-оболонці Jupyter, яка:

- формує вибірку;
- стратифікує дані за декількома критеріями;
- використовує нейронну мережу;
- застосовує колаборативний фільтр.

Як результат роботи цієї програми, отримали збільшення конверсії в 8 разів.



## 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

На сьогоднішній день все більше стартапів перетворюють ідеї пов'язані із нейромережевими технологіями у працюючі бізнес-моделі. Інтерес інвесторів до стартапів, що працюють із нейронними мережами зростає з кожним роком. Все більше і більше з'являється прикладів успішного застосування штучного інтелекту у найрізноманітніших галузях діяльності людини, значно зростає кількість корпорацій що впроваджують системи прогнозування із застосуванням нейромереж у свою операційну діяльність.

В тому числі ми можемо застосовувати нейронні мережі в напрямку роздрібної торгівлі.

### 4.1. Опис ідеї проекту.

Результативність застосування традиційних методів прогнозування фінансових активів, наприклад акцій, облігацій, та валюти, які вільно продаються і купуються на біржах, можна назвати недостатньою для потреб сучасного ринку. Це пов'язано із тим, що інвестиції на фондовому ринку тісно пов'язані із Інтернетом і залежні від інформаційного середовища. Для підвищення точності прогнозування доцільно застосувати таку модель, що не тільки базується на кореляціях факторів та особливостях часового ряду, а й тісно пов'язана з декількома джерелами даних.

Сучасні системи прогнозування не враховують комплексно кількісні та якісні фактори, що впливають на зміну курсів акцій або враховують у векторному вигляді, який обмежує можливості представлення вхідних даних до нейромережі. Такі

системи прогнозування зазвичай мають точність 53-58% вірного прогнозу. Цієї точності замало для того, щоб використовувати такі системи як повноцінний інструмент для економічного аналізу та прогнозування. Отже, ідеєю стартап проекту є створення і розповсюдження системи прогнозування фінансових ринків із більшою ніж у конкурентів точністю прогнозування тренду (табл. 4.1).

Таблиця 4.1 Опис ідеї стартап-проекту

Напрямки застосування	Вигоди для користувача
1.Управління взаємовідносин с клієнтом	Аналоги (на загальному ринку) відсутні. Застосування математичного апарату з метою покращення якості взаємовідносин з клієнтом та збільшення прибутків.

Застосування зрозумілої математичної моделі спрощує реалізацію.

Метою аналізу техніко-економічних переваг є чітке виокремлення технічних і маркетингових особливостей розробленого продукту:

#### 4.2. Технологічний аудит проекту.

Метою технічного аудиту є визначення переліку технологій, за допомогою який реалізована система і їх аналіз. Визначення технологічної здійсненності ідеї проекту передбачає аналіз таких складових (таблиця 4.2):

- 1) за допомогою якої технологією розроблена система згідно ідеї проекту;

- 2) чи існують у відкритому доступі ці технології, чи їх потрібно додатково розробляти або купувати;
- 3) чи має доступ розробник до описаних технологій.

Таблиця 4.2. Технологічна здійсненність ідеї проекту

№п/п	Ідея проекту	Технології реалізації	Наявність технологій	Доступність технологій
		нейронні мережі, колаборативні фільтри		
Для взаємодії з даними використовується Microsoft SQL, що в подальшому опрацьовується в мові програмування Python				

За результатами аналізу можна зробити висновок щодо можливості технологічної реалізації проекту. Технологічним шляхом реалізації проекту було обрано мову програмування Python через її доступність та безкоштовність.

#### 4.3. Аналіз ринкових можливостей запуску стартап-проекту.

Метою аналізу ринкових можливостей є виявлення та дослідження обмежень, можливостей та розрахунок конкретних показників реалізації розробленої системи прогнозування фінансових ринків як стартап-проекту.

Спочатку було проведено аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (таблиця 4.3).

Таблиця 4.3 - Попередня характеристика потенційного ринку стартап проекту

№ п/п	Показники стану ринку(найменування)	Характеристика
1	Кількість систем-конкурентів на ринку, од	0
2	Загальний обсяг продаж, грн/ум.од	0
3	Динаміка ринку	Повільний приріст
4	Наявність обмежень для входу (вказати характер обмежень)	Нема
5	Специфічні вимоги до стандартизації та сертифікації	Нема
6	Середнє значення рентабельності в галузі(або по ринку), %	50%

Проаналізувавши результати, можна зробити висновок що проект є придатним до інвестицій, оскільки вкладені кошти будуть повернені за 2 роки.

За результатами порівняння що було наведене у таблиці 4.3 було зроблено висновок, що ринок є придатним для розповсюдження системи як стартап-проекту.

Після цього були досліджені потенційні категорії клієнтів, їх особливості та затверджено перелік вимог до системи для кожної категорії клієнтів (табл. 4.4)

Таблиця 4.4 — Характеристика потенційних клієнтів стартап-проект

№ п/п	Потреба, що формує ринок системи CRM	Цільова аудиторія та потенційні клієнти	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги до споживачів товару
1	Нейронна мережа з застосуванням колаборативної фільтрації	Компанії що мають клієнтську базу	Цільові групи клієнтів мають різні доходи і потенційний прибуток від цієї системи буде різний	Оптимізовані та автоматизовані алгоритми, кроссплатформенність

Після дослідження потенційних категорій клієнтів було проведено дослідження ринкового середовища: складено таблиці факторів, що сприяють реалізації розробленої системи як стартап-проекту, та факторів, що йому заважають (табл. 4.5, 4.6).

Таблиця 4.5 - Фактори загроз

№п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Конкуренція	Вихід на ринок систем з кращими Характеристиками та моделлю надання послуг	Вийти на ринок зосередивши увагу на власній перевазі системи. Покращення якісних характеристик системи прогнозування. Обрати цільову аудиторію
2	Зміна потреб користувачів	Клієнту потрібна буде система з більшою точністю прогнозування і новими функціями	Передбачити можливість розширення системи та підвищення точності прогнозування

Таблиця 4.6 - Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Конкуренція	Відсутність аналогів на українському ринку для вітчизняного користувача	Адаптація системи до особливостей потреб на українському ринку
2	Поява альтернативних методів моделювання	Нові методи моделювання, більш легкі в освоєнні праці	Розширення можливостей, максимальне спрощення

Надалі було проведено дослідження пропозиції: визначили загальні характеристик конкуренції на ринку (таблиця 4.7).

Таблиця 4.7 — Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому виражена дана характеристика	Вплив на діяльність компанії (можливі дії для підвищення конкурентоспроможності)
1. Вказати тип конкуренції - олігополія	На ринку присутні декілька постачальників-конкурентів, але їх товар дещо відрізняється від нашої системи.	Підтримка якості системи, безперервний розвиток, покращення, вдосконалення, оновлення та підтримка.
2. За рівнем конкурентної боротьби - міжнародний	Компанії-конкуренти з інших країн	Створити основу системи прогнозування таким чином, щоб можна було легко локалізувати її для використання у інших країнах.
3. Загалузовою ознакою - внутрішньогалузева	Система може застосовуватися в одній галузі, але її різних сферах.	Постійне вдосконалення системи прогнозування, що не має прив'язки до сфери, але має до галузі

Продовження табл. 4.7

Особливості конкурентного середовища	В чому виражена дана характеристика	Вплив на діяльність компанії (можливі дії для підвищення конкурентоспроможності)
4. Конкуренція за видами товарів: - товарно-видова	Конкуренція між видами систем прогнозування, їх особливостями.	Створити систему прогнозування, враховуючи недоліки конкурентів
5. За характером конкурентних переваг - цінова	Покращення процесу створення програмного продукту, мінімізація витрат на оновлення, застосування безперервної інтеграції	Використання відкритих та дешевших технологій для побудови системи в порівнянні з системами-конкурентами, але тільки якщо ці технології відповідають необхідним критеріям якості.
6. За інтенсивністю - не марочна	Бренд присутній, але його роль незначна	Реклама, участь у конференціях, семінарах, виставках.

Було проведено аналіз конкуренції у галузі за моделлю М. Портера (табл. 4.8).



Таблиця 4.8 — Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Навести перелік безпосередніх конкурентів	Визначити бар'єри входження в ринок	Визначити фактори сили постачальників	Визначити фактори сили клієнтів	Фактори загроз з боку товарів-замінників
	Neural Builder	Наявність вже існуючих рішень	-	Якість системи та її підтримка оновлення	Більш відомий розробник, що підтримує свою систему
Висновки:	На даний момент немає конкурентів на українському ринку	Вихід на український ринок буде легшим через відсутність конкуренції	-	Вимоги клієнтів такі, як зручний інтерфейс, якість Програмного продукту	Випустити систему, що буде не гірше, ніж у конкурента, але мати кращу точність розрахунків

За результатами порівняльного дослідження що наведене у табл. 4.8 було зроблено висновок про доцільність виходу на український та міжнародні ринки.

На основі аналізу ситуації на ринках, проведеного в табл. 4.8, а також із урахуванням характеристик розробленої системи, вимог потенційних клієнтів до товару (табл. 4.4) та особливостей ринкового середовища (таблиці 4.5, 4.6) досліджується та визначається перелік факторів конкурентоспроможності розробленої системи прогнозування фінансових ринків.

За визначеними факторами конкурентоспроможності проведено аналіз сильних та слабких сторін стартап-проекту (табл. 4.9).

Таблиця 4.9 — Порівняльний аналіз сильних та слабких сторін проекту

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг систем-конкурентів у порівнянні з розробленою системою прогнозування						
			-3	-2	-1	0	+1	+2	+3
1	Ціна	15					+		
2	Кросплатформність	20			+				
3	Орієнтованість на кінцевого споживача	7					+		

Кінцевим кроком маркетингового дослідження можливостей реалізації системи прогнозування фінансових ринків як стартап-проекту є побудова SWOT-аналізу (матриці сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (таблиця 4.10) на основі описаних конкурентних та маркетингових загроз та можливостей, та сильних і слабких сторін (таблиця 4.9). Список маркетингових загроз та можливостей було складено на основі дослідження

факторів загроз та факторів можливостей ринкової ситуації. Маркетингові загрози та можливості є наслідками (прогнозованими результатами) впливу ринкових факторів.

Таблиця 4.10 — SWOT-аналіз стартап-проекту

Сильні сторони:  Ціна, орієнтованість на кінцевого споживача, кросплатформеність	Слабкі сторони:  Складність розповсюджувати продукцію за кордоном.
Можливості:  Відсутність конкуренції на українському ринку	Загрози:  Зміна основних потреб клієнтів, при відсутності конкуренції необхідно підтримувати інтерес аудиторії до продукту

За результатами SWOT-аналізу було сформовано альтернативи ринкової стратегії (перелік заходів) для виведення розробленої системи як стартап-проекту на український та міжнародні ринки та розрахований оптимальний час їх ринкової реалізації з урахуванням потенційних розробок конкурентів, що можуть бути виведені на ринок (див. таблицю 4.8 аналіз потенційних конкурентів). Запропоновані альтернативи були проаналізовані з точки зору часу реалізації та ймовірності отримання ресурсів (таблиця 4.11).

Таблиця 4.11 — Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний Комплекс заходів) ринкової Поведінки	Приблизна ймовірність отримання ресурсів	Приблизні строки реалізації
1	Безкоштовне розповсюдження Обмеженої версії створеного програмного продукту	85%	12 місяців
2	Створення Програмної системи з подальшим платним розповсюдженням (продаж платної ліцензії)	45%	12 місяців

Після аналізу було обрано альтернативу №2.

#### 4.4. Розроблення ринкової стратегії проекту.

Розроблення ринкової стратегії першим етапом передбачає визначення стратегії охоплення ринку: дослідження цільових груп потенційних клієнтів, які приведені в таблиці 4.12.

Таблиця 4.12. Вибір цільових груп потенційних споживачів

№п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Компанії що займаються фінансовою аналітикою	Готові	Необхідно	Висока	Середня

Визначена цільова група клієнтів: Компанії що займаються фінансовою аналітикою

Дослідивши потенційні групи клієнтів було визначено цільові категорію, для яких буде пропонуватися розроблена система, та визначено стратегію охоплення ринку - стратегію диференційованого маркетингу.

Для роботи в обраних сегментах ринку сформовано базову стратегію розвитку (таблиця 4.13).

Таблиця 4.13 — Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку стартап-проекту	Обрана альтернатива розвитку стартап-проекту	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1		Визначити потреби сучасного ринку для кожної з груп.	Цінова політика, універсальність продукту.	Стратегія Диференціації

Наступним кроком обрано стратегію конкурентної поведінки (таблиця 4.14).

Таблиця 4.14 — Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект першою подібною системою на ринку?	Чи буде розробник системи шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде розробник копіювати властивості товару конкурента	Стратегія конкурентної поведінки*
1	Ні	Шукати Нових	Ні	Заняття конкурентної ніші

За результатами аналізу вимог клієнтів визначених категорій до розробника стартап-проекту та до продукту (див. таблицю 4.5), а також в залежності від обраної

базової стратегії розвитку (таблиця 4.13) та стратегії конкурентної поведінки (таблиця 4.6) розроблено стратегію позиціонування (таблиця 4.15), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Таблиця 4.15 — Визначення стратегії позиціонування

№ п/п	Вимоги до системи з боку потенційних клієнтів	Основна стратегія розвитку	Ключові конкуренто спроможні позиції власного стартап-проекту	Вибір основних асоціацій
1	Проста побудова моделі прогнозування, довгий час навчання нейромережі, проте вища точність прогнозу та інтуїтивно зрозумілий інтерфейс, докладне керівництво користувача	Стратегія диференціації	Позиція на основі порівняння фірми з товарами конкурентів; Відмінні особливості споживача	Економія часу; Зручність застосування; Практичність

За результатами дослідження була сформована система рішень щодо ринкової поведінки стартап-компанії, яка визначає напрями роботи стартап-компанії українському та міжнародному ринках.

#### 4.5. Розроблення маркетингової програми стартап-проекту

Визначено маркетингову концепцію системи використання нейронної мережі, яку буде купувати споживач. Основна концепція продукту — письмовий опис фізичних та програмних характеристик системи, які сприймаються споживачем, і набору вигід, які він обіцяє певній групі споживачів.

За результатами дослідження було сформовано трирівневу маркетингову модель системи: ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання:

- 1-й рівень вирішує питання щодо того, засобом вирішення якої потреби і / або проблеми буде система, яка її основна вигода;
- 2-й рівень являє рішення того, як буде реалізована система в реальному ринковому середовищі. Рівень включає в себе якість, властивості, дизайн, упаковку, ціну;
- 3-й рівень визначає додаткові послуги та переваги для клієнта системи, що створюються на основі товару за задумом і товару в реальному виконанні (гарантії якості , доставка, умови оплати та ін).

Після цього визначимо цінові межі, якими необхідно керуватись при встановленні ціни на систему, яке включає аналіз ціни на товари-аналоги або товари субститутути, а також аналіз рівня доходів цільової категорії клієнтів (таблиця 4.16). Аналіз проведено експертним методом.



Таблиця 4.16 — Визначення меж встановлення ціни

№ п/п	Приблизна вилка вартості товарів- замінників	Приблизна вилка вартості товарів- аналогів	Приблизний рівень доходів цільової групи клієнтів	Верхня та нижня межі вартості системи
1	800-3000\$	800-3000\$	3000\$+	200-700\$

Після цього визначимо оптимальну систему збуту, за допомогою якої буде розповсюджуватися система як сатрап-проект.(таблиця 4.17)

Таблиця 4.17 — Формування системи збуту

№ п/п	Особливості ринкової поведінки потенційних клієнтів	Функції збуту, які має виконувати постачальник системи	Глибина каналу збуту	Оптимальний канал збуту
1	Цільові клієнти – компанії, які бажають впровадити у своїй роботі сучасні засоби допоможуть прогнозувати фінансові ринки із кращою точністю	Побудова прямих контактів із потенційними клієнтами і їх підтримка. формування попиту і стимулювання продажів.	Один (від виробника одразу споживачу)	Прямий канал збуту до клієнта, мінімізувати збутові витрати. Розвиток нових маркетингових концепцій.

Фінальною складовою маркетингової програми є визначення концепції маркетингових комунікацій.

Результатом дослідження стала робоча ринкова програма, що включає в себе концепції системи, збуту, просування та попереднє дослідження ціноутворення на продукт, базується на цінностях та потребах категорій покупців, конкурентні переваги системи, стан та динаміку маркетингового середовища, в межах якого впроваджено системи прогнозування фінансових ринків як стартап-проект, та відповідну обрану альтернативу ринкової поведінки.

#### 4.6. Висновки до розділу

В цьому розділі було проведено аналіз розробленої системи максимізації вартості клієнта (CVM) у якості стартап-проекту. Варто зауважити, що проект має можливість ринкової комерціалізації, через те, що ринок потребує якісного та інноваційного продукту для прогнозування поведінки клієнта.

Результатом роботи є розроблений стартап-проект, план виходу на ринки програмного забезпечення та маркетингова стратегія. Розроблений стартап-проект доцільно застосувати при комерціалізації розробки.

Висока конкуренція зумовлює необхідність виваженого підходу до просування продукту. Проте кількість учасників фінансових ринків зростає з кожним днем, що зумовлює хороші перспективи для комерціалізації системи. Для впровадження ринкової реалізації проекту була обрана стратегія, яка включає розробку програмної системи із класичною моделлю ліцензування за певну плату.

Можна сказати, що подальший розвиток проекту є доцільним, оскільки кількість потенційних користувачів системи зростає кожного року.

## ВИСНОВКИ

В рамках даної роботи було присвячено увагу максимізації вартості клієнта. Для досягнення цієї мети було сконцентровано увагу на наступному:

1. Сегментація клієнтів.
2. Висунення гіпотез.
3. Аналіз та збільшення значення LTV (lifetime value), час життя клієнта.
4. Аналіз маркетингових дій та їх подальше покращення та оптимізація.

Для вирішення поставлених задач, були застосовані наступні інструменти:

1. Проведення кампаній на стратифікованих пілотних та контрольних групах. Що були сегментовані за рядом критеріїв:
  - а) розмір чека;
  - б) частота відвідування;
  - в) ціль відвідування.
2. Висунення та перевірка гіпотез, оцінка їх ефективності за допомогою конверсії оціненої на інкрементальному заробітку.
3. Застосування колаборативних фільтрів та нейронної мережі з метою підвищення показників конверсії (більш влучним застосуванням наших пропозицій).

Вирішення задачі реалізоване у вигляді програми на мові програмування Python в веб-оболонці Jupyter.

Результати магістерської дисертації: за допомогою застосування нейронної мережі та колаборативного фільтра, вдалось збільшити значення конверсії в понад 8 разів.

## ПЕРЕЛІК ПОСИЛАНЬ

1. "CRM History: The Evolution Of Better Customer Service". URL: [https://www.streetdirectory.com/travel\\_guide/124130/enterprise\\_information\\_systems/crm\\_history\\_the\\_evolution\\_of\\_better\\_customer\\_service.html](https://www.streetdirectory.com/travel_guide/124130/enterprise_information_systems/crm_history_the_evolution_of_better_customer_service.html)
2. Bardicchia, Marco. Digital CRM: Strategies and Emerging Trends: Building Customer Relationship in the Digital Era. Independently published: 2020. 106 p.
3. "Management Tools - Customer Relationship Management - Bain & Company". URL: [www.bain.com](http://www.bain.com).
4. Neural network. URL: [https://en.wikipedia.org/wiki/Neural\\_network](https://en.wikipedia.org/wiki/Neural_network)
5. Колаборативна фільтрація. URL: <https://uk.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BB%D0%B0%D0%B1%D0%BE%D1%80%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%B0%D1%84%D1%96%D0%BB%D1%8C%D1%82%D1%80%D0%B0%D1%86%D1%96%D1%8F>
6. John M. Carroll. Human-Computer Interaction in the New Millennium. Boston: Addison-Wesley Professional, 2001. 752 p.
7. Nancy R. Tague. The Quality Toolbox, Second Edition. Wisconsin: ASQ Quality Press Milwaukee, 2005. 584 p.
8. Савчук О.В., Кривенко К.С. Інтелектуальний аналіз діагностичної інформації електро- радіокомпонентів в умовах невизначеності: Інтелектуальний аналіз інформації: зб. праць. Київ: Просвіта, 2013. С. 211-217
9. Терехов В.А., Ефимов Д.В., Тюкин И.Ю. Нейросетевые системы управления: учеб. пособие для вузов. Москва: Высшая школа, 2002. 183 с.
10. Горбань А.Н. Нейронные сети на персональном компьютере. Новосибирск: Наука, 1996. 276 с.

11. Лисе А.А., Степанов М.В. Нейронные сети и нейрокомпьютеры. Учеб. пособие. СПб.: ГЭТУ., 2009. 64 с.
12. Царегородцев В.Г. Перспективы распараллеливания программ нейросетевого анализа и обработки данных: материалы с III Всерос. конф. «Математика, информатика, управление – 2008». Иркутск, 2014. С. 110-117
13. Рутковская Д.А., Пилиньский М.О., Рутковский Л.Р. Нейронные сети, генетические алгоритмы и нечёткие системы: пер. с польск. И.Д. Рудинского. Москва: Горячая линия Телеком, 2006. 452 с.
14. Саймон Хайкин Нейронные сети: полный курс, 2-е издание: пер. с англ. Москва: Издательский дом "Вильямс", 2006. 1045 с.
15. Осовский С. Нейронные сети для обработки информации: пер. с польского. Москва: Финансы и статистика, 2002. 344 с.
16. Круглов В.В. Нечёткая логика и искусственные нейронные сети. Москва: Физматлит, 2001. 224 с.
17. Дюк В.В, Самойленко А.В. Data Mining: учебный курс, СПб: Питер. 2001. 368 с.
18. Нейромережеве відображення дійсності. URL:  
[http://studies.in.ua/mpd\\_seminar/1313-neyrmerezh.html](http://studies.in.ua/mpd_seminar/1313-neyrmerezh.html)
19. Jonathan L. Herlocker, Joseph A Konstan, Loren G Terveen, John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*. 2004. Vol. 22, No. 1. P. 5-53

## ДОДАТОК А ЛІСТИНГ ПРОГРАММИ

```

import sys
import os
import imp
import pandas as pd
from sklearn.model_selection import train_test_split
import CampaignTestingModule as ctm

import pyodbc
import sqlalchemy
import urllib

def pipeline(tableWithAudience,
             sprintNumber,
             pg_size=10000,
             cg_size=10000,
             requiredNumberOfGroups=100):

    db_conn, engine = connectToDB()
    members = getMembersFromDB(tableWithAudience, db_conn)
    members = addFieldsForStratification(members)

    groups = split_into_groups(members, pg_size, cg_size, requiredNumberOfGroups)
    groups = enumerate_groups(groups)

    campaign = pd.concat(groups)
    campaign = addIsPGField(campaign)
    campaign = campaign[campaign['group_id'] <= requiredNumberOfGroups]

    return campaign

def getMembersFromDB(tableWithAudience,
                     db_conn):

    sql = getSqlForMembersRequest(tableWithAudience)
    members = pd.read_sql(sql, db_conn)

    return members

def getSqlForMembersRequest(tableWithAudience):

    sql = f"""
    select *

```

```

from Marketing.cvm.{tableWithAudience}
"""

return sql

def addFieldsForStratification(members):

    members['avg_cheque_segment'] = pd.qcut(members['avg_cheque'], 5, duplicates='drop').cat.codes
    members['buys_count_segment'] = pd.qcut(members['buys_count'], 5, duplicates='drop').cat.codes

    return members

def split_into_groups(members_clean,
                      pg_size,
                      cg_size,
                      requiredNumberOfGroups):

    campaign = members_clean.copy().sample(frac=1)
    groups = []
    for campaign_num in range(requiredNumberOfGroups):

        if campaign_num == 0:
            group_size = cg_size
        else:
            group_size = pg_size

        campaign, group = train_test_split(campaign, test_size=group_size,
                                           random_state=42, stratify=campaign[['channel', 'avg_cheque_segment',
                                           'buys_count_segment']])
        groups.append(group)

    return groups

def enumerate_groups(groups):
    for i, group in enumerate(groups):
        group['group_id'] = i + 1
    return groups

def addIsPGField(campaign):
    """
    Добавляет поле is_pg:
    1 - если 'group_id' <> 1
    0 - если 'group_id' == 1

```

```
'''
```

```
def caseForIsPG(row):  
    if row['group_id'] == 1:  
        return 0  
    else:  
        return 1
```

```
campaign['is_pg'] = campaign.apply(caseForIsPG, axis=1)
```

```
return campaign
```